

L Number	Hits	Search Text	DB	Time stamp
1	148358	DNA or polynucleotide or RNA adj2 shuffling or reco,b\$	USPAT; US-PGPUB; DERWENT	2003/09/11 09:55
2	209004	DNA or polynucleotide or RNA adj2 shuffling or recomb\$	USPAT; US-PGPUB; DERWENT	2003/09/11 09:55
3	26815	(DNA or polynucleotide or RNA adj2 shuffling or recomb\$) and single-strand\$	USPAT; US-PGPUB; DERWENT	2003/09/11 09:55
4	4773	((DNA or polynucleotide or RNA adj2 shuffling or recomb\$) and single-strand\$) and exonuclease	USPAT; US-PGPUB; DERWENT	2003/09/11 09:56
5	3826	((((DNA or polynucleotide or RNA adj2 shuffling or recomb\$) and single-strand\$) and exonuclease) and PCR	USPAT; US-PGPUB; DERWENT	2003/09/11 09:56
6	3721	(((((DNA or polynucleotide or RNA adj2 shuffling or recomb\$) and single-strand\$) and exonuclease) and PCR) and alter\$	USPAT; US-PGPUB; DERWENT	2003/09/11 09:57
7	3548	((((((DNA or polynucleotide or RNA adj2 shuffling or recomb\$) and single-strand\$) and exonuclease) and PCR) and alter\$) and primers	USPAT; US-PGPUB; DERWENT	2003/09/11 09:57
8	259	((((((((DNA or polynucleotide or RNA adj2 shuffling or recomb\$) and single-strand\$) and exonuclease) and PCR) and alter\$) and primers) and BAL31	USPAT; US-PGPUB; DERWENT	2003/09/11 09:58

# Optimization of DNA shuffling for high fidelity recombination

Huimin Zhao and Frances H. Arnold\*

Division of Chemistry and Chemical Engineering 210-41, California Institute of Technology, Pasadena, CA 91125, USA

Received December 19, 1996; Accepted January 27, 1997

## ABSTRACT

**A convenient 'DNA shuffling' protocol for random recombination of homologous genes *in vitro* with a very low rate of associated point mutagenesis (0.05%) is described. In addition, the mutagenesis rate can be controlled over a wide range by the inclusion of  $Mn^{2+}$  or  $Mg^{2+}$  during DNase I digestion, by choice of DNA polymerase used during gene reassembly as well as how the genes are prepared for shuffling (PCR amplification versus restriction enzyme digestion of plasmid DNA). These protocols should be useful for *in vitro* protein evolution, for DNA based computing and for structure-function studies of evolutionarily related genes.**

The method of DNA shuffling, or 'sexual PCR', is used to recombine homologous DNA sequences during *in vitro* molecular evolution (1,2). While randomly recombining the DNA sequences, the technique also introduces new point mutations at a relatively high rate (0.7%; 3). Though these point mutations may provide useful diversity for some *in vitro* evolution applications, they are problematic for others, especially when the mutation rate is this high. Much lower mutagenesis rates are desired, for example, during the *in vitro* evolution of long genes or whole operons (4), during recombination of beneficial mutations already identified previously (5), for DNA-based computing, or when the method is used to differentiate adaptive from neutral or deleterious mutations in evolutionarily-related sequences (6).

In order to optimize the DNA shuffling technique with regard to fidelity, we have attempted to minimize the number of point mutations introduced in each step. Here we describe a convenient DNA shuffling protocol which randomly recombines genes with a very low rate of associated point mutagenesis. In addition, we show how the mutagenesis rate associated with DNA shuffling can be controlled over a practically useful range by appropriate changes in the protocol.

DNA shuffling consists of four steps: (i) preparation of genes to be shuffled, (ii) fragmentation with DNase I, (iii) reassembly by thermocycling in the presence of a DNA polymerase, and (iv) amplification of reassembled products by a conventional PCR. Point mutations may be generated during each of these steps. Lorimer and Pastan reported that use of  $Mn^{2+}$  instead of  $Mg^{2+}$  during the DNase I fragmentation step improves the fidelity of DNA shuffling ~3-fold (7). Our protocols include this improvement. Conventional PCR with *Taq* polymerase is usually used to prepare the genes to be shuffled (step 1) as well as to amplify the

reassembled products (step 4) (1,7). However, the fidelity of *Taq* polymerase is the lowest among commercially available thermostable DNA polymerases. In fact, between 33 and 98% of the amplification products will contain mutation(s) when a 1 kb fragment is amplified for 20 effective cycles (one million-fold amplification) using *Taq* polymerase. Extensive studies have revealed that fidelity during PCR depends on the specific conditions and DNA polymerase used (8-10). Avoiding PCR where possible and using higher fidelity DNA polymerase during amplification and reassembly should further reduce the point mutagenesis rate associated with DNA shuffling.

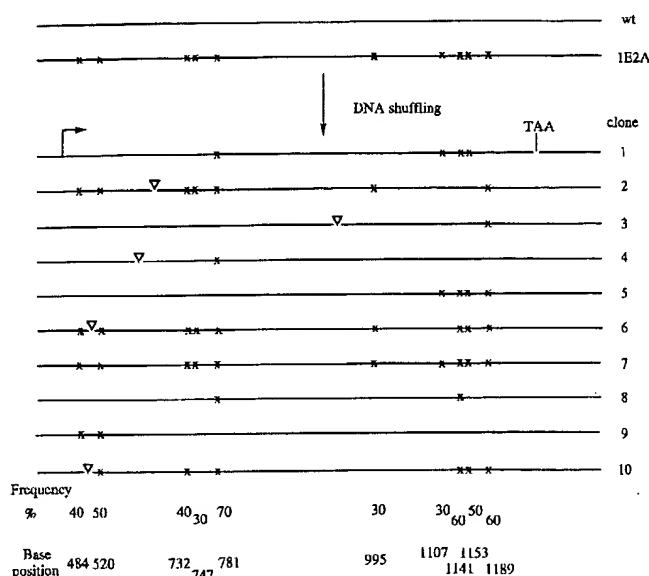
Wild-type subtilisin E and its thermostable mutant 1E2A genes were randomly recombined by DNA shuffling using the conditions summarized in Table 1. Gene 1E2A, obtained by directed evolution of wild-type subtilisin E, differs by 10 base changes (6; Fig. 1). The enzyme encoded by this gene retains wild-type activity. The ~1 kb fragments encoding mature subtilisin E from residue -15 (from the prosequence) to the C-terminus (including 113 nt after the stop codon) were obtained by restriction digestion of plasmid DNA and purified from a 0.8% agarose gel using the QIAEX II gel extraction kit (QIAGEN, Chatsworth, CA). After DNA shuffling, the gene library was amplified in *E. coli* HB101 and transferred into *B. subtilis* DB428 competent cells for expression and screening, as described elsewhere (6). Screening for protease activity was carried out at 37 °C in 96-well plates using suc-Ala-Ala-Pro-Phe-*p*-nitroanilide (0.2 mM) as substrate, as described previously (11). All PCR reactions were done on a MJ Research (Watertown, MA) PTC200 thermocycler. Sequencing was done on an ABI 373 DNA Sequencing System using the Dye Terminator Cycle Sequencing kit (Perkin-Elmer, Branchburg, NJ).

The frequency of clones exhibiting  $\geq 10\%$  of wild-type subtilisin activity was used as a convenient index of the fidelity of DNA shuffling under different conditions. Table 1 shows that only 20% of the clones retained subtilisin activity after DNA shuffling using the protocol of Lorimer and Pastan (conditions A) (7). Replacing *Taq* with *Pwo* polymerase in the reassembly step (conditions B) increased this to 46%.

Improvements in fidelity could also be achieved by preparing the genes directly from plasmids by restriction enzyme digestion (conditions C-E). This, together with PCR amplification using a 1:1 mixture of *Taq* and *Pfu*, yielded 58% active clones (conditions C). Finally, the use of proofreading enzymes (*Pfu* or *Pwo*) in the reassembly step dramatically increased the frequency of active clones to as high as 95% (conditions E). These data are consistent with the known fidelities of the various DNA polymerases.

To determine the mutagenesis rate of our optimized DNA shuffling protocol (conditions E) as well as to analyze the efficiency

\*To whom correspondence should be addressed. Tel: +1 818 395 4162; Fax: +1 818 568 8743; Email: frances@cheme.caltech.edu



**Figure 1.** Results of sequencing genes from 10 randomly-selected unscreened clones from DNA shuffled library. Lines represent 986 bp of subtilisin E gene including 45 nt of its prosequence, the entire mature sequence and 113 nt after the stop codon. Crosses indicate positions of mutations from 1E2A, while triangles indicate positions of new point mutations introduced during the DNA shuffling procedure.

**Table 1.** Frequency of active clones obtained after DNA shuffling under different conditions

Step Condition	1 Gene preparation	2 DNase I digestion <sup>a</sup>	3 Reassembly	4 PCR amplification	Active clones <sup>b</sup> %
A	PCR / <i>Taq</i>	Mn <sup>2+</sup>	<i>Taq</i>	<i>Taq</i>	20
B	PCR / <i>Taq</i>	Mn <sup>2+</sup>	<i>Pwo</i>	<i>Taq</i>	46
C	plasmid digestion	Mn <sup>2+</sup>	<i>Taq</i>	<i>Taq:Pfu</i> (1:1)	58
D	plasmid digestion	Mn <sup>2+</sup>	<i>Pwo</i>	<i>Taq:Pfu</i> (1:1)	87
E	plasmid digestion	Mn <sup>2+</sup>	<i>Pfu</i>	<i>Taq:Pfu</i> (1:1)	95

<sup>a</sup>As suggested in ref. 7.

<sup>b</sup>Clones exhibiting >10% wild-type subtilisin E activity.

of recombination, 10 unscreened clones were selected randomly for sequencing. The plasmids from these clones were purified and the inserts were sequenced in both directions. Comparison of these 10 sequences (Fig. 1) shows that all except clone 7 result from different recombination events. (Clone 7 is the intact 1E2A parent sequence.) The frequency of occurrence of a particular point mutation from parent 1E2A in the shuffled genes ranged from 30 to 70%, fluctuating around the expected value of 50%. No pair of point mutations was found to be inseparable, even those as little as 12 bp apart. However, a certain degree of linkage of nearby mutations is apparent. In 9860 total bases sequenced, no insertion or deletions (frameshifts) were found. The overall mutagenic rate for this protocol is only 0.05% (5/9860), which is the lowest rate thus far reported for DNA shuffling.

### High-fidelity DNA shuffling protocol

**Preparation of genes to be shuffled.** About 10 µg plasmids containing wild-type subtilisin E and the thermostable mutant

1E2A gene were digested at 37°C for 1 h with *Nde*I and *Bam*HI (30 U each) in 50 µl 1× buffer B (Boehringer Mannheim, Indianapolis, IN). Fragments of ~1 kb were purified from 0.8% preparative agarose gels using QIAEX II gel extraction kit (QIAGEN, Chatsworth, CA). The DNA fragments were dissolved in 10 mM Tris-HCl (pH 7.4). The DNA concentrations were estimated, and the fragments were mixed 1:1 for a total of ~2 µg.

**DNase I digestion in the presence of Mn<sup>2+</sup>.** The mixture was diluted to 45 µl in 10 mM Tris-HCl (pH 7.4) and 5 µl 10× digestion buffer (500 mM Tris-HCl pH 7.4, 100 mM MnCl<sub>2</sub>) was added. This mixture was equilibrated at 15°C for 5 min on a thermocycler before 0.30 U DNase I (10 U/µl; Boehringer Mannheim) was added. The digestion was done at 15°C and terminated after 2 min by heating at 90°C for 10 min. That the fragments were <50 bp was confirmed on a 2% agarose gel before purification on a Centri-Sep column (Princeton Separations, Inc., Adelphia, NJ).

**Fragment reassembly.** Spin-column purified fragments (10 µl) were added to 10 µl 2× PCR premix [5-fold diluted cloned *Pfu* buffer, 0.4 mM each dNTP, 0.06 U/µl cloned *Pfu* polymerase (Stratagene, La Jolla, CA)]. The reaction mixture was overlaid with 30 µl mineral oil. PCR program: 3 min 96°C followed by 40 cycles of 1 min 94°C, 1 min 55°C, 1 min + 5 s/cycle 72°C, followed by 7 min at 72°C.

**PCR amplification of reassembled products.** One microlitre of this reaction was used as template in a 25-cycle PCR reaction. PCR conditions (100 µl final volume): 30 µmol each primer, 1× *Taq* buffer, 0.2 mM each dNTP and 2.5 U *Taq/Pfu* (1:1) mixture. PCR program: 2 min 96°C, 10 cycles of 30 s 94°C, 30 s 55°C, 45 s 72°C, followed by another 14 cycles of 30 s 94°C, 30 s 55°C, 45 s + 20 s/cycle 72°C, and finally 7 min 72°C. This program gives a single band at the correct size.

We have developed a high-fidelity DNA shuffling protocol with mutagenic rate of only 0.05%. The mutagenic rate can be controlled over a wide range, 0.05–0.7%, by the inclusion of Mn<sup>2+</sup> or Mg<sup>2+</sup>, by the choice of DNA polymerase and/or using restriction enzyme digestion, as indicated in Table 1. Further improvements in fidelity could be achieved by reducing the PCR cycle number in the reassembly step and using *Pfu* only in the final amplification step. The current high fidelity protocol has been used successfully to distinguish the functional and non-functional mutations in an evolved gene by random recombination and sequence analysis of genes conferring the evolved phenotype (6). This approach should prove extremely useful for structure–function studies of homologous genes.

### REFERENCES

- 1 Stemmer, W.P.C. (1994) *Nature* **370**, 389–391.
- 2 Cramer, A., Whitehorn, E.A., Tate, E. and Stemmer, W.P.C. (1995) *Nature Biotech.* **14**, 315–319.
- 3 Stemmer, W.P.C. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 10747–10751.
- 4 Moore, J.C. and Arnold, F.H. (1996) *Nature Biotech.* **14**, 458–467.
- 5 Moore, J.C. and Arnold, F.H. (1997) *Adv. Biochem. Engng.* **58**, in press.
- 6 Zhao, H. and Arnold, F.H., submitted.
- 7 Lorimer, I.A.J. and Pastan, I. (1995) *Nucleic Acids Res.* **23**, 3067–3068.
- 8 Kunkel, T.A., Loeb, L.A. and Goodman, M.F. (1984) *J. Biol. Chem.* **259**, 1539–1545.
- 9 Keohavong, P. and Thilly, W.G. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9253–9257.
- 10 Eckert, K.A. and Kunkel, T.A. (1990) *Nucleic Acids Res.* **18**, 3739–3744.
- 11 Chen, K. and Arnold, F.H. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 5618–5622.